

APJ Abdul Kalam Technological University
 First Semester M.Tech Degree Examination December 2016
 Ernakulam II Cluster
 COMPUTER SCIENCE AND ENGINEERING
 Specialization: COMPUTER SCIENCE AND ENGINEERING

05CS 6005 – DATA MINING AND WAREHOUSING

Time : 3 hrs.

Max. Marks: 60

I

- a) Scores for 6 students are: **10 20 30 40 50 60**

Find the median and quartiles for this distribution. Also draw the box plot.

(3 marks)

- b) Discuss the role of data cube aggregation in data reduction process?

(3 marks)

- c) Consider the following transactions

<u>Transaction-id</u>	<u>Items bought</u>
10	A, B, D
20	A, C, D
30	A, D, E
40	B, E, F
50	B, C, D, E, F

Let $\text{min_sup} = 60\%$ and $\text{min_conf} = 50\%$. Find out the frequent itemsets using Apriori algorithm. (6 marks)

II

- a) What are the measures taken to avoid overfitting in classification? (4 marks)

- b) Classify an unknown tuple (Red Domestic SUV) using Naïve Bayesian Classifier.

(8 marks)

No	Color	Type	Origin	Stolen ?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes

8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No

III

- a) What is a data ware house? Explain the three tier architecture of a data warehouse with a neat diagram. (8 Marks)
- b) Explain the working of K Means and K Medoids Partitioning methods. (10 Marks)

OR

IV

- a) What is OLAP? Explain the typical OLAP operations involved in the multidimensional data model using an example. (8 Marks)
- b) Explain the working of the following clustering algorithms with an example
- i. BIRCH (5 Marks)
 - ii. CLIQUE (5 Marks)

V

- a) Which are the different types of outliers? Explain the different types of statistical methods for outlier detection. (12 Marks)
- b) How will you perform clustering based outlier detection? Discuss the clustering based approach for identifying outliers in small clusters. (6 Marks)

OR

VI

- a) Explain the classification of proximity based outlier detection methods. (12 Marks)
- b) How outlier detection can be used for high dimensional data applications? (6 Marks)